r = rnorm(100,mean=100,sd=10)

> length(r[r>=(mean(r)-sd(r))&r<=(mean(r)+sd(r))])

[1] 65 ≈ 68%

> length(r[r>=(mean(r)-2\*sd(r))&r<=(mean(r)+2\*sd(r))])

[1] 96≈ 95%

> length(r[r>=(mean(r)-3\*sd(r))&r<=(mean(r)+3\*sd(r))])

[1] 99≈ 99.7%

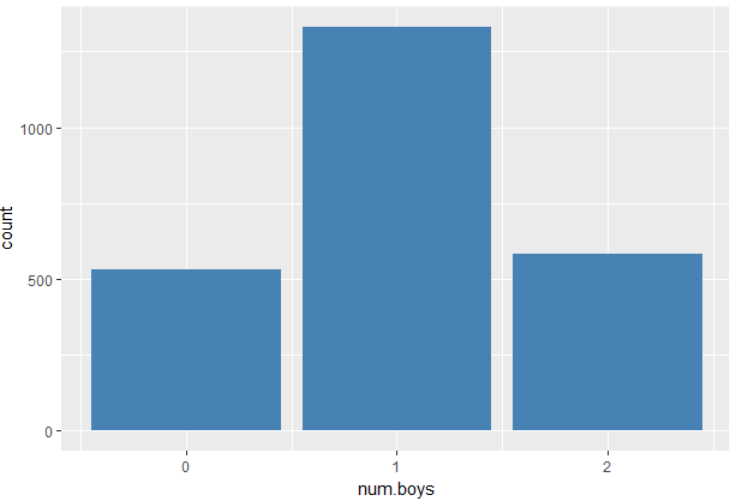
install.packages(“abd”)

library(abd)  
library(ggplot2)

data = TwoKids

> data['num.boys'] = as.character(data$num.boys)

ggplot(data, aes(x = num.boys, y = count)) + geom\_bar(stat='identity',fill='steelblue')



> data = read.csv("C:\\Users\\Administrator\\Desktop\\statistical computing\\HW03\\airmay.csv")  
> data1 = na.omit(data)

> summary(data1[,c('X1','X2','X3','Y')])

X1 X2 X3 Y

Min. : 8.00 Min. : 5.700 Min. :57.00 Min. : 1.00

1st Qu.: 74.75 1st Qu.: 9.575 1st Qu.:61.00 1st Qu.: 11.00

Median :206.50 Median :11.500 Median :66.00 Median : 18.00

Mean :182.04 Mean :11.504 Mean :66.46 Mean : 24.12

3rd Qu.:292.25 3rd Qu.:12.750 3rd Qu.:69.75 3rd Qu.: 32.50

Max. :334.00 Max. :20.100 Max. :81.00 Max. :115.00

> par(mfrow = c(2,2))

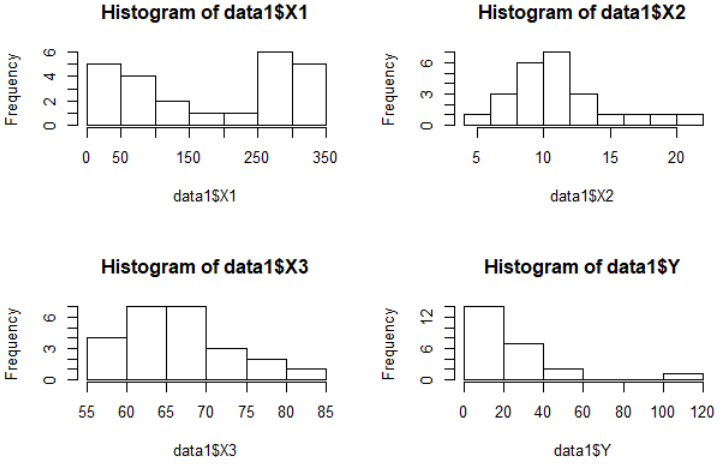
> hist(data1$X1)

> hist(data1$X2)

> hist(data1$X3)

> hist(data1$X4)

> hist(data1$Y)



> t.test(data1$Y,conf.level = 0.9)$conf.int

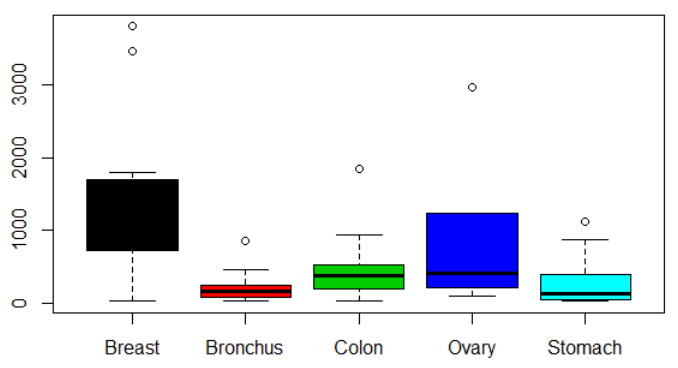
[1] 16.11852 32.13148

attr(,"conf.level")

[1] 0.9

> data = read.table("C:\\Users\\Administrator\\Desktop\\statistical computing\\HW03\\question4.txt",header = T)

> boxplot(data$Survival~factor(data$Organ),col=c(1,2,3,4,5))



> attach(data)

> tapply(Survival,Organ,summary)

$Breast

Min. 1st Qu. Median Mean 3rd Qu. Max.

24 723 1166 1396 1692 3808

$Bronchus

Min. 1st Qu. Median Mean 3rd Qu. Max.

20.0 72.0 155.0 211.6 245.0 859.0

$Colon

Min. 1st Qu. Median Mean 3rd Qu. Max.

20.0 189.0 372.0 457.4 519.0 1843.0

$Ovary

Min. 1st Qu. Median Mean 3rd Qu. Max.

89.0 239.8 406.0 884.3 1039.5 2970.0

$Stomach

Min. 1st Qu. Median Mean 3rd Qu. Max.

25 46 124 286 396 1112

> data = PlantGrowth

> nrow(data)

[1] 30

> tapply(weight,group,mean)

ctrl trt1 trt2

5.032 4.661 5.526

> t.test(data[data[, "group"] == "trt1", "weight"],data[data[, "group"] == "trt2", "weight"],alt = 'two.sided')

Welch Two Sample t-test

data: data[data[, "group"] == "trt1", "weight"] and data[data[, "group"] == "trt2", "weight"]

t = -3.0101, df = 14.104, p-value = 0.009298

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-1.4809144 -0.2490856

sample estimates:

mean of x mean of y

4.661 5.526

p-value <0.05, reject H0, there is difference between trt1 group and trt2 group

> install.packages("RSADBE")

library(RSADBE)

­> data(Gasoline)  
> names(Gasoline)

[1] "y" "x1" "x2" "x3" "x4" "x5" "x6" "x7" "x8" "x9" "x10" "x11"

>t.test(Gasoline[Gasoline[,'x11']=='M','y'],Gasoline[Gasoline[,'x11']=='A','y'],alternative = "greater")

Welch Two Sample t-test

data: Gasoline[Gasoline[, "x11"] == "M", "y"] and Gasoline[Gasoline[, "x11"] == "A", "y"]

t = 4.1695, df = 7.1911, p-value = 0.001975

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

5.933054 Inf

sample estimates:

mean of x mean of y

28.42429 17.58611

p-value < 0.05, reject null hypothesis.

> library("UsingR")

> t.test(babies$dage,babies$age,alt="greater")

Welch Two Sample t-test

data: babies$dage and babies$age

t = 11.067, df = 2301.5, p-value < 2.2e-16

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

2.865266 Inf

sample estimates:

mean of x mean of y

30.73706 27.37136

p-value < 0.05, reject H0, do not have enough evidence to prove that the difference is not greater than 0.

> data = read.csv("C:\\Users\\Administrator\\Desktop\\statistical computing\\HW03\\Noshow.csv")

> dim(data)

[1] 110527 14

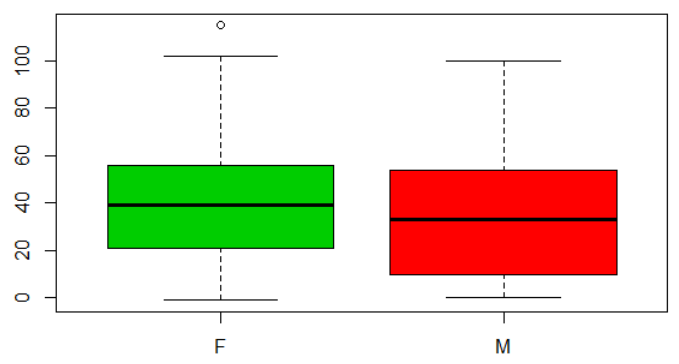
> names(data)

[1] "PatientId" "AppointmentID" "Gender" "ScheduledDay" "AppointmentDay"

[6] "Age" "Neighbourhood" "Scholarship" "Hipertension" "Diabetes"

[11] "Alcoholism" "Handcap" "SMS\_received" "No.show"

> boxplot(data$Age~factor(data$Gender),col = c(3,2))



> attach(data)

> xtabs(~Gender+No.show)

No.show

Gender No Yes

F 57246 14594

M 30962 7725

> prop.test(x=c(14594,7725),n=c(71840,38687),alt='greater')

2-sample test for equality of proportions with continuity correction

data: c(14594, 7725) out of c(71840, 38687)

X-squared = 1.8534, df = 1, p-value = 0.08669

alternative hypothesis: greater

95 percent confidence interval:

-0.0007094951 1.0000000000

sample estimates:

prop 1 prop 2

0.2031459 0.1996795

p-value > 0.05, not enough evidence to reject null hypothesis.

> xtabs(~SMS\_received+No.show)

No.show

SMS\_received No Yes

0 62510 12535

1 25698 9784

> prop.test(x=c(9784,12535),n=c(35482,75045),alt='greater')

2-sample test for equality of proportions with continuity correction

data: c(9784, 12535) out of c(35482, 75045)

X-squared = 1766.1, df = 1, p-value < 2.2e-16

alternative hypothesis: greater

95 percent confidence interval:

0.1041922 1.0000000

sample estimates:

prop 1 prop 2

0.2757454 0.1670331

p-value < 0.05, reject null hypothesis.

>t.test(data[data$Gender=="F","Age"],data[data$Gender=="M","Age"],alt="greater")

Welch Two Sample t-test

data: data[data$Gender == "F", "Age"] and data[data$Gender == "M", "Age"]

t = 34.561, df = 72834, p-value < 2.2e-16

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

4.911678 Inf

sample estimates:

mean of x mean of y

38.89399 33.73686

P-value < 0.05, reject null hypothesis.